

You Spot the Fake: When Literacy Replaces Accountability

Weekly Analysis — <https://ainews.social>

Somewhere in the middle of this week's most widely shared French primer on synthetic media, after the alarm has been rung in full — artificial intelligence, the guide announces, has broken the pact of trust that once bound images to reality — the reader is handed a to-do list. Look at the eyes: are they blinking too little, or too much? Study the lighting; count the fingers; watch the hairline where the wig of pixels meets the skull. The crisis is civilizational. The remedy is a scavenger hunt. That gap — between the size of the problem named and the smallness of the task assigned — is the most honest thing about the week's literacy discourse, because it tells you exactly where the repair work has been outsourced to. It has been outsourced to you [7].

This is the move to watch. When a technology dissolves the evidentiary floor under public life, someone has to be made responsible for the dissolving. The literacy artifacts that flooded the week — deepfake-spotting guides, cybersecurity explainers, reskilling forecasts, advice columns for anxious parents — converge on a single, quietly radical answer: the responsible party is the viewer. Not the firm that trained the model, not the platform that ranked the forgery into your feed, not the state that decided your suspicion was a threat to public order. You. The instruction to "spot the glitch" arrives dressed as empowerment, and it is worth taking seriously as empowerment — but it is also, read against the same week's evidence, the final quality-control station on an assembly line that literacy claims to oppose. UNESCO's own media-literacy handbook concedes the point almost in passing, noting that detecting deepfakes "sometimes require expertise and particular competencies similar to those used in forensic science" [1]. We have redefined the ordinary citizen as an unpaid forensic examiner and called it education.

What I want to map here is the conceptual muddle inside the word "literacy" itself, because the muddle is not accidental. There are at least two literacies competing under one name. One teaches you to inspect the artifact — to become a better consumer of a compromised information supply, a discerning shopper in a market of fakes. The other teaches you to interrogate the supply itself — who manufactures it, who profits from its manufacture, and who has an interest in you

[7] Deepfakes et désinformation : comment l'IA menace la confiance

[1] UNESCO Think Critically Click Wisely

never asking. This week the first literacy won, decisively and almost everywhere. The purpose of this essay is to explain why that victory is not the neutral, obviously-good thing it presents itself as, and what a literacy worthy of a democracy would have to teach instead.

The Pact, and Who Broke It

Start with the framing device that organizes nearly every artifact: the broken pact. It is a genuinely useful phrase. For most of the photographic century, an image carried a rebuttable presumption of contact with the real — someone stood where the camera stood, and light bounced off something that existed. Generative systems retire that presumption. The French guide is right to name this as a rupture in trust rather than a mere uptick in fakery [7]. Alvin Toffler saw the shape of it decades early: information becoming “not only hyperkinetic, but also ephemeral,” until “facts, information, knowledge, and history could become increasingly perishable” and a “post-fact society” ceases to be a slogan and becomes a description [1].

But notice what happens the instant the pact is declared broken. The declaration is passive — the pact “broke,” the way a glass breaks, as if by accident or weather. No agent appears. And because no agent appears, no agent can be held to account, and the only available site of intervention becomes the endpoint: the human eye. The guide’s grammar does the ideological work before any advice is given. A pact that broke on its own requires no one to answer for it; it requires only that survivors adapt. Kate Crawford’s insistence that we refuse “the notion that artificial intelligence is a purely technical domain,” and instead see it as “technical and social practices, institutions and infrastructures, politics and culture,” is precisely the refusal the broken-pact framing forecloses [1]. The pact did not break. Specific companies built specific systems that generate specific forgeries, and specific platforms distribute them because distribution is profitable. To say “the pact broke” is to launder all of that into meteorology.

This is where the two literacies first split. A consumer-protection literacy accepts the passive voice and gets to work on coping. A civic literacy treats the passive voice as the first thing to be analyzed — asks who is served by the disappearance of the agent, and puts the agent back into the sentence. The week’s flagship materials chose coping, and once you have chosen coping, the curriculum writes itself: perception, forensics, vigilance. The trouble is that this curriculum has a documented expiry date, and everyone teaching it knows.

[7] Deepfakes et désinformation : comment l’IA menace la confiance

[1] After shock

[1] The Atlas of AI - Power, Politics, and the Planetary Costs

A Checklist With an Expiry Date

Here is the load-bearing weakness of perceptual-forensics literacy, and it is not subtle. The tells it teaches are the exact defects that model vendors are paid to eliminate. UNESCO's handbook, to its credit, prints the current forensic wisdom plainly: watch whether "persons blinking too much or too little," and remember that deepfakes "often fail to fully represent the natural physics of a scene... natural physics of lighting" [1]. Every one of those sentences is a specification document for the next training run. Blinking artifacts were a 2019 problem; the physics of lighting is a 2022 problem. To teach a citizen in 2026 to hunt for them is to train them, with great diligence, for a model generation that has already been decommissioned. You are issuing spotter's guides to an aircraft that no longer flies.

This is what makes the literacy-as-forensics program structurally, rather than incidentally, doomed. It is not that the checklist is imperfect and will improve. It is that the checklist is a moving target moving in the wrong direction — every published tell is an errand handed to the adversary's engineers, who close the gap faster than any curriculum can be revised. Daniel Susskind, writing this week as a father trying to explain the technology to his own children, lands near this honesty when he warns parents that the reassuring intuitions we hand our kids about spotting fakery are already obsolete, and that the harder task is teaching judgment about sources and incentives rather than surface cues [14]. The advice that scales is not "look closer." The advice that scales is "the pixels will win the perception war, so stop fighting it there."

Consider the modality where the tells have already collapsed entirely: voice. The synthetic "family emergency" call — a cloned voice of your child or grandchild, panicked, needing money now — has no hairline to inspect and no blink rate to count [4]. The entire visual-forensics apparatus is inapplicable, and what remains is precisely the civic-critical faculty that perceptual literacy neglects: knowledge of how scams are structured, of what urgency-plus-secrecy demands signify, of the social engineering that makes the clone effective regardless of its acoustic fidelity. Educators working the fraud beat have begun to say this outright — that literacy in the age of scams has to be about the architecture of manipulation, not the artifact's surface [20]. And that architecture is not perceptual at all. A controlled experiment run on 10,101 people this week found that AI-driven persuasion could measurably shift participants' financial choices — not by deceiving their eyes with a better fake, but by working their reasoning, their biases, their trust [17]. You cannot spot your way out of being

[1] UNESCO Think Critically Click Wisely

[14] I'm a father of three who studies the impact of artificial intelligence: this is what parents need to know about AI

[4] Arnaques au clonage vocal par IA : le faux appel d'« urgence familiale »

[20] Teaching AI Literacy in the Age of Scams: A Conversation with Dr Gabriel Aguilar

[17] Manipulation par l'IA : testée sur 10 101 personnes, elle change leurs choix d'argent

persuaded. The glitch was never the threat.

The Taught Threat and the Actual Threat

Follow the money and the mismatch becomes glaring. The literacy discourse teaches the deepfake as spectacle — the uncanny video, the impossible celebrity, the face that isn't there. The threat it teaches you to fear is a perceptual event. But the threat the same week actually documents is an industrial one, and the two are separated by an order of magnitude in both cost and design.

The touchstone case, invoked whenever institutions want to dramatize the stakes, is the ARUP fraud: a finance worker at the engineering firm wired roughly twenty-five million dollars after a video call populated entirely by deepfaked colleagues, a chief financial officer and staff who were all synthetic. It is a genuinely instructive event, and the cybersecurity literature has adopted it as its founding parable, treating literacy as, in one framing, "the new frontline" of enterprise defense [2]. But notice how the parable is bounded. It is a corporate emergency, a story about protecting balance sheets, and it locates the failure in an individual employee's insufficient suspicion. The moral is: be more vigilant, verify through a second channel, don't trust the video call. Consumer protection for the enterprise.

Now set beside it the threats that do not fit the corporate frame. In Colombia, in the run-up to a presidential second round, digital clones of the politician Iván Cepeda were manufactured and circulated to spread disinformation into a live electoral contest — not to defraud a treasury but to bend a vote [6]. This is not a twenty-five-million-dollar operation requiring a live synthetic video call; the economics of automated influence have collapsed to something closer to the cost of a modest monthly subscription. The Reuters Institute has spent the year arguing, against the deepfake-panic grain, that the more consequential danger is not the dazzling one-off forgery but the cheap, high-volume, humans-in-the-loop production of tailored falsehood at scale [12]. The taught threat is a jewel heist. The actual threat is a factory.

And the factory is already running. Investigators this week documented 232 Facebook accounts pumping AI-generated "health slop" — fabricated medical advice engineered to sell supplements — to a combined audience of eighty-one million followers [10]. None of that content asks to survive a forensic inspection of its lighting. It works by volume and plausibility, by flooding the zone until the true and the false cost the same to produce and the reader's discernment simply

[2] AI Literacy as the New Frontline of Cybersecurity

[6] Clones digitales de Iván Cepeda difundieron desinformaciones antes de la segunda vuelta presidencial en Colombia

[12] How AI-generated disinformation might impact this year's elections and how journalists should report

[10] From Doctors to Healers: 232 Facebook Accounts Publish Health 'AI Slop' to Push Supplements, Guides and Misinformation to Their 81 Million Followers

exhausts. Meanwhile a subtler front has opened: rather than fooling humans directly, disinformation operations now aim to poison the AI systems humans increasingly ask for answers, seeding the web so that when you query a chatbot about a candidate or a war you are served the laundered propaganda as neutral summary [19]. The Institute for Strategic Dialogue tested exactly this exposure, evaluating how AI assistants answer election questions and finding real gaps in accuracy, sourcing, and language coverage [5]. Here the "spot the fake" curriculum is not merely obsolete; it is category-inapplicable. There is no artifact to inspect. The forgery has moved upstream of your eyes, into the machine you were told to trust more than your eyes.

To be fair to the sober end of this literature, the panic itself can be overstated, and the Knight First Amendment Institute has usefully cautioned against treating every election as an AI-apocalypse [8]. But the corrective cuts toward my point, not away from it: the Reuters Institute's companion argument is that you should "worry more about humans" than about autonomous AI systems, because the damage is done by human institutions deploying these tools with intent [11]. If the danger is human intent operating through cheap infrastructure, then a literacy that trains only the retina has aimed itself at the one place the danger isn't.

When Doubt Becomes Disloyalty

Now the argument turns, because there is a second silence in the week's literacy discourse even more consequential than the expired checklist. Perceptual literacy assumes a benign environment: a citizen freely exercising critical judgment, a state that welcomes the exercise, a public sphere that rewards doubt. The week's own evidence describes the opposite — an environment in which the governing powers treat critical doubt itself as a security problem. And a literacy that teaches you to doubt images while remaining silent about who is permitted to doubt, and about whom, has taught you to be careful in a room whose exits are watched.

Look at Indonesia, where the new AI regulations were built, an analysis this week argues, to leave state repression conspicuously unchecked — the rules discipline citizens' uses while exempting the government's own [15]. Look at Italy, where Giorgia Meloni's decree expanding AI facial recognition drew alarm from experts who saw a surveillance capacity growing faster than any accountability for it [9]. In both cases the AI environment the "critically literate" citizen is meant to navigate is not a neutral commons; it is a regulated,

[19] Russia's disinformation war has a new target: the AI that answers your questions

[5] Chatbots and the Ballot Box: Evaluating Accuracy, Sourcing and Language Gaps in AI Answers to Election Questions

[8] Don't Panic (Yet): Assessing the Evidence and Discourse Around Generative AI and Elections

[11] Generative AI and elections: why you should worry more about humans than AI systems

[15] Indonesia's AI regulations leave state repression unchecked

[9] El decreto de Meloni de reconocimiento facial con IA genera preocupación entre los expertos

surveilled, and at points hostile space, engineered so that the state's fakery and the state's watching are the least accountable elements in it. The literacy guides say nothing about this. They cannot, because to name it would be to admit that the problem is governance, not perception.

The Colombian case sharpens the knife. When voters raised concerns about the integrity of the electoral information environment — an environment demonstrably contaminated by synthetic clones of a candidate — a European electoral observation figure was reported to have waved those concerns away as mere *narrativas*, stories, the paranoid fiction of sore losers [6]. Sit with the structure of that. The same week's literacy discourse tells the citizen: be suspicious, the images may be fake. And the same week's authority tells the citizen who acts on that suspicion: your suspicion is itself a fabrication, a narrative, a threat to legitimacy. The citizen is caught in a pincer. Doubt is mandated as a consumer skill and pathologized as a civic act. You are trained to distrust the pixels and shamed for distrusting the power that governs the pixels.

[6] Clones digitales de Iván Cepeda difundieron desinformaciones antes de la segunda vuelta presidencial en Colombia

This is where Chomsky and Herman's old analysis of the manufactured consensus becomes newly literal. Their model held that the information environment is filtered by ownership, advertising, sourcing, and flak — the disciplining pushback that punishes deviation from acceptable framing [1]. The Colombian "narrativas" dismissal is flak in its purest form: an institutional mechanism for making the doubting citizen the problem. A literacy that maps only the artifact and never the filter leaves the citizen fully equipped to identify a bad JPEG and fully disarmed to identify the machinery that decided which doubts are permissible. Safiya Noble's formulation, cited in the metaliteracy literature, names the deeper design flaw: algorithmic oppression "is not just a glitch in the system but, rather, is fundamental to the operating system of the web" [1]. The whole rhetorical genius of "spot the glitch" is that it insists the trouble is a glitch — a local defect in an otherwise trustworthy system — precisely when the trouble is the operating system.

[1] Manufacturing Consent

[1] MetaLiteracyIACW

Literacy Folded Into Compliance

There is a third literacy in the week's discourse, and it arrives wearing the language of opportunity rather than danger: the reskilling forecast. This is the register in which literacy is defined not as civic capacity nor even as consumer vigilance but as human-capital maintenance — the continuous obligation of the worker to remain employable against

a moving technological frontier. IBM’s framing is representative and worth reading closely, because it presents AI literacy as, above all, the closing of a “skills gap,” a deficit in the workforce to be remediated so the workforce stays productive [3]. The projections that circulate in this register — that some enormous fraction of the global workforce, a majority in many national estimates, will need to retrain before 2030 — arrive with an air of natural law, like weather, like the broken pact.

And they carry the same buried passive voice. Who decided the workforce needs retraining? Who wrote the curriculum? Who pays for it — the worker, on unpaid evenings, or the firm that automated the old job? The reskilling literature almost never asks. It treats the necessity as given and the syllabus as neutral, and in doing so it performs the exact operation that empties literacy of its critical content. A literacy defined as compliance with a curriculum you did not choose, to remain useful to an economy that did not consult you, has kept the word and discarded the thing. Crawford’s insistence that AI is “institutions and infrastructures, politics and culture” applies with full force here: the reskilling mandate is a political demand disguised as a technical fact, a redistribution of adjustment costs onto the least powerful party, narrated as personal development [1].

The cybersecurity framing feeds the same logic from a different direction. Calling AI literacy “the new frontline” enlists the citizen-worker as a volunteer soldier in a perimeter defense, responsible for repelling attacks that the institutions above them profit from making cheap [2]. The frontline metaphor sounds like empowerment and functions like conscription. It is the ARUP employee again, now generalized to the whole population: the failure is always at the endpoint, the remedy is always more vigilance from the person with the least power to change the system, and the system itself — the vendors selling both the generative weapon and the detection shield, the platforms monetizing the flood — remains outside the frame of what literacy is permitted to examine.

It is worth naming the counter-move that broke from this pattern, because it clarifies the alternative by contrast. When New York’s mayor and schools chancellor imposed a broad moratorium on generative AI in schools, the striking thing was the shift in the locus of responsibility: instead of instructing children to become better forensic examiners of an environment saturated with synthetic content, the authority intervened on the environment itself [18]. One can argue about whether a moratorium is wise or durable — reasonable people will disagree — but it is at least a policy that treats the information environment as governable rather than as weather to be endured with a checklist. It answers the question the forensic guides refuse to ask:

[3] Alfabetización en IA: cerrar la brecha de habilidades de la IA

[1] The Atlas of AI - Power, Politics, and the Planetary Costs

[2] AI Literacy as the New Frontline of Cybersecurity

[18] Mayor Mamdani and Chancellor Samuels Put Students First with Nation’s Broadest Generative AI Moratorium in Schools

who is responsible for the conditions, and can those conditions be changed by anyone other than the person at the endpoint?

The Literacy Worth the Name

So what would the second literacy — the civic one, the one that lost this week — actually teach? Not less about deepfakes, but a different relation to them. It would teach the perceptual tells as a footnote with an explicit expiry date, honestly labeled as a stopgap for the current model generation, so that no citizen mistakes a scavenger hunt for security. And it would spend its real energy on the questions the artifact-focused curriculum cannot ask: who owns the systems that generate synthetic media, who ranks it into your attention, who is exempted from the rules, and who has decided that your suspicion is disloyalty. The distinction is between a literacy that makes you a sharper consumer of a compromised supply and a literacy that makes you a citizen capable of asking who governs the supply.

There are fragments of this better literacy in the week’s evidence, and they point the way. The most promising research direction abandons the spot-the-fake model entirely in favor of inoculation — building generalizable resistance to the techniques of manipulation before exposure, so that the citizen recognizes the emotional and rhetorical machinery of a false claim regardless of whether its surface is a perfect forgery [21]. That is a literacy of structure rather than surface, and it survives the next model release because it never depended on the last one’s defects. The election-integrity work that maps how to detect and guard against synthetic political content treats the problem as one of institutional defense and public infrastructure, not lone-viewer vigilance [13]. And the more thoughtful education scholarship this week reframed generative AI through what it called a framework of agency — asking not merely how learners can identify AI outputs but how they can retain authorship, judgment, and power within an AI-saturated environment [16]. Agency is the right word, because it is the thing the forensic checklist quietly takes away. To be handed a list of tells is to be told your job is to inspect, not to govern.

The stakes of the choice between these two literacies are not academic, and they are not confined to media literacy as a niche. Every other domain in which AI now operates — the workplace being reskilled, the ballot being contested, the state doing its watching — depends on which literacy prevails, because literacy is the faculty through which citizens either understand those domains or are managed within them. Toffler’s warning about a post-fact society was not

[21] Towards Generalizable AI-Assisted Misinformation Inoculation

[13] How to Detect and Guard Against Deceptive AI-Generated Election Information

[16] La IA generativa en la educación: un marco de agencia

that facts would disappear but that they would become "perishable," and a perishable-fact world is survivable only by a public that can ask where facts come from and who has an interest in spoiling them [1]. A public trained only to squint at hairlines will not ask. It will inspect the pixels dutifully, pass the quality-control check, and hand the certified product back to the pipeline that made it — never having been taught that the pipeline was the thing to interrogate.

[1] After shock

That is the trap in "you spot the fake." It flatters the citizen as the last line of defense while quietly making the citizen the last quality-control station of the very apparatus producing the fakes. The alarm about the broken pact is real, and the guides are right to sound it. But the moment they shrink the response to the individual retina, they have decided — on your behalf, without telling you — that accountability is off the table and that literacy will do the work accountability declined to do. Until literacy learns to ask who broke the pact, who profits from its brokenness, and who has ruled your doubt out of order, "spot the glitch" is not the citizen's defense against the pipeline. It is the pipeline's final, unpaid inspector, wearing the badge of empowerment and calling it education.

References

1. After shock
2. AI Literacy as the New Frontline of Cybersecurity
3. Alfabetización en IA: cerrar la brecha de habilidades de la IA
4. Arnaques au clonage vocal par IA : le faux appel d'« urgence familiale »
5. Chatbots and the Ballot Box: Evaluating Accuracy, Sourcing and Language Gaps in AI Answers to Election Questions
6. Clones digitales de Iván Cepeda difundieron desinformaciones antes de la segunda vuelta presidencial en Colombia
7. Deepfakes et désinformation : comment l'IA menace la confiance
8. Don't Panic (Yet): Assessing the Evidence and Discourse Around Generative AI and Elections
9. El decreto de Meloni de reconocimiento facial con IA genera preocupación entre los expertos
10. From Doctors to Healers: 232 Facebook Accounts Publish Health 'AI Slop' to Push Supplements, Guides and Misinformation to Their 81 Million Followers

11. Generative AI and elections: why you should worry more about humans than AI systems
12. How AI-generated disinformation might impact this year's elections and how journalists should report
13. How to Detect and Guard Against Deceptive AI-Generated Election Information
14. I'm a father of three who studies the impact of artificial intelligence: this is what parents need to know about AI
15. Indonesia's AI regulations leave state repression unchecked
16. La IA generativa en la educación: un marco de agencia
17. Manipulation par l'IA : testée sur 10 101 personnes, elle change leurs choix d'argent
18. Mayor Mamdani and Chancellor Samuels Put Students First with Nation's Broadest Generative AI Moratorium in Schools
19. Russia's disinformation war has a new target: the AI that answers your questions
20. Teaching AI Literacy in the Age of Scams: A Conversation with Dr Gabriel Aguilar
21. Towards Generalizable AI-Assisted Misinformation Inoculation